Homework 6

Please complete all exercises in Jupyter notebook and turned in a knitted PDF with output included. Be sure to tune your Spark code with appropriate configuration parameters. During the execution of your code be sure to use the Spark UI to track your jobs and use available information from the UI to help you tune your jobs. If you don’t need all of the available memory on your PC to run the job be sure to tune it down.

1. Setup a Spark Structured Streaming session to stream in the JSON data in the /data directory of the lab.
2. Calculate the following statistics by Country:
   1. Average temperature in Celsius
   2. Number of records
3. Write the resultant Spark dataframe as a CSV to a SEPARATE directory.
4. Turn in your Notebook as PDF along with your CSV.